Leveraging Free AI and ML Resources for Drug Optimization

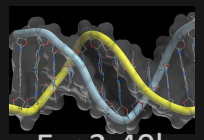
Boutique R&D Services from Switzerland

Oliv Eidam

CDD Community Meeting :: Feb 27, 2025

Free AI / ML Resources Leveraging my Work

- AlphaFold2
- ML models and scripts on git
- Public datasets (ChEMBL and specialized sets)
- LLM models for biology (ESM-2, EVO-2)
- Antibody models (Ablang2)
- ChemProp (molecular representation)
- RDKit, DataWarrior (free tools for cheminfo)
- ChatGPT



Evo2-40b

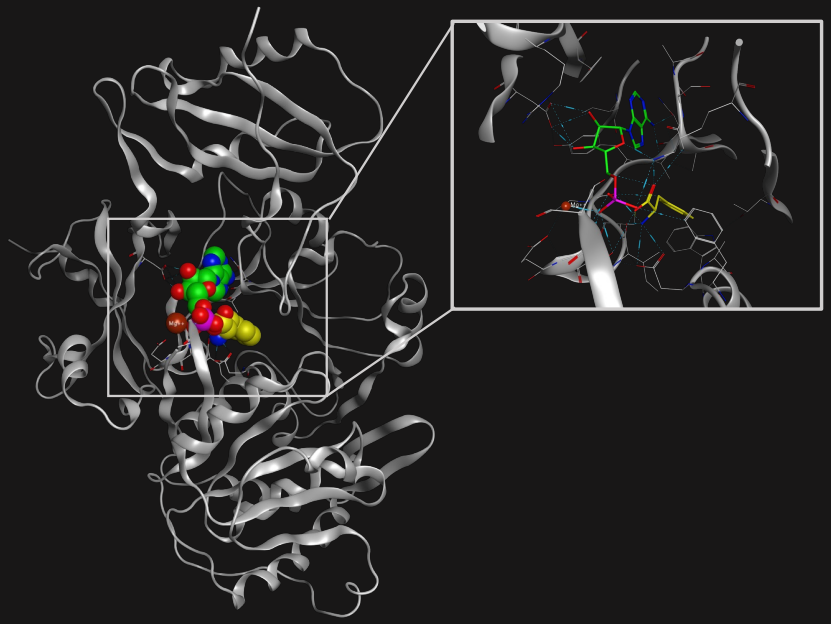
Oxford Protein Informatics Group (OPIG)

Next: 3 Case Studies

Image sources: https://alphafold.ebi.ac.uk, https://github.com, https://about.gitlab.com, https://www.rdkit.org , https://openmolecules.org/datawarrior, https://build.nvidia.com/arc/evo2-40b, https://opig.stats.ox.ac.uk, https://chatgpt.com

Study 1: AlphaFold2 for Homology Modeling

- **Challenge**: Model a target with 38% sequence identity
- **Benefits**: Fast, deep alignments, fewer human errors, available
- **Luck**: AF2 matched closed, inhibitor-bound form which was desired for inhibitor modeling
- **Guidance**: Used 2D images of unpublished X-rays
- **Limitation**: Co-factors, metals, inhibitors required separate modeling



Structure depicts X-ray of enzyme with ADP (green), Mg^{2+} (brown sphere) and a ligand (yellow). This is not the modeled target.

Study 2: ChemProp for Property Predictions

- **ChemProp**: Graph Neural Network (GNN) representation of molecules works well for property prediction of small molecules[1]
- **Peptide Membrane permeability** : ChemProp equally predictive (RMSE 0.75) and easier to use than alternative deep learning method (RMSE 0.77)[2]
- **Data and scripts**: see [3] and [4]

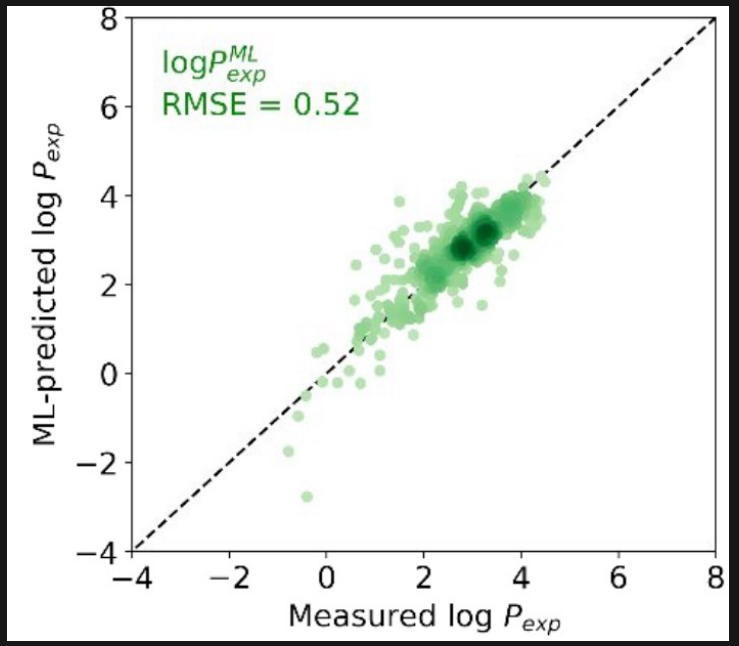
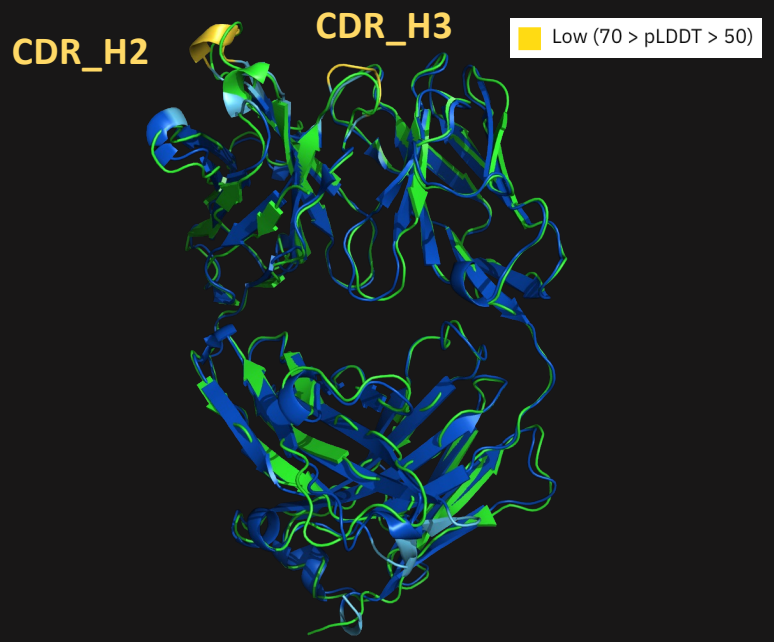


Figure S14 from [1] depicting ChemProp-predicted log P against experimentally measured log P for 425 test set molecules from public AZ dataset (2538 mols).

*[1] Clemens Isert, Jimmy C. Kromann, Nikolaus Stiefl, Gisbert Schneider, Richard A. Lewis, ACS Omega. **2023** Jan 4;8(2):2046-2056. Machine Learning for Fast, Quantum Mechanics-Based Approximation of Drug Lipophilicity*
*[2] Jianan Li, Keisuke Yanagisawa, Yutaka Akiyama, Brief Bioinform. **2024** Jul 25;25(5). CycPeptMP: enhancing membrane permeability prediction of cyclic peptides with multi-level molecular features and data augmentation*
[3] https://github.com/cisert/rescoss_logp_ml
[4] CycPetpMPDB: http://cycpeptmpdb.com

Study 3: Antibody Modeling Boosted by AI

- **AF2-multimer:** Good for scaffolds, less confident in long CDRs[1]
- **Ablang2**: best freely available antibody sequence model[2]
- **ESM-2**: Boosts affinity, stability, reduces poly-specificity and immunogenicity[3,4]
- **Evo-2**: Largest bio-LLM[4]



Superposition of AF2-model (blue) on Fab X-ray (green) shows little error (RMSD 1.16 Å) and perfect hetero-dimer of heavy and light chains. Only CDR_H2 and CDR_H3 exhibit low confidence (70 > pLDDT > 50).

[1] My favorite Jupyter notebook implementation of AF2: https://colab.research.google.com/github/sokrypton/ColabFold/
[2] https://github.com/oxpig/AbLang2
*[3] Lin Z. et al, Science. **2023** Mar 17;379(6637):1123-1130. Evolutionary-scale prediction of atomic-level protein structure with a language model*
*[4] Brian Hie et al, Nat Biotechnol. **2024** Feb;42(2):275-283. Efficient evolution of human antibodies from general protein language models.*
*[5] Brian Hie and co-workers, bioRxiv, **Feb 19, 2025**. DOI. Genome modeling and design across all domains of life with Evo 2*

AI in Daily Use: Benefits & Limitations. Next?

AI in Daily Use

- AI improves my emails and coding
- Study 1: AlphaFold2 reduces modeling errors, but ligands require separate modeling
- Study 2: ChemProp elevates peptide permeability prediction
- Study 3: AI models are powerful in antibody modeling but struggle with variable loops
- All Studies: Open-source code & data are valuable

Next Steps: Let's Discuss!

- AlphaFold*: Small molecule integration?
- New ML molecular representations?
- Merging physics with ML for predictions?

Thank you!